AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1002/2014JD021696

Key Points:
- A high-resolution model captures variabilities of temperature and precipitation
- The simulation is shown to add value above coarse-resolution data
- Simulation sensitivities to model setup and physics schemes are explored

Supporting Information:
- Readme
- Figures S1–S5 and Tables S1 and S2

Correspondence to:
V. R. Kotamarthi,
vrkotamarthi@anl.gov

Citation:
Wang, J., and V. R. Kotamarthi (2014), Downscaling with a nested regional climate model in near-surface fields over the contiguous United States, *J. Geophys. Res. Atmos.*, *119*, 8778–8797, doi:10.1002/2014JD021696.

Received 25 FEB 2014
Accepted 9 JUL 2014
Accepted article online 13 JUL 2014
Published online 28 JUL 2014

Downscaling with a nested regional climate model in near-surface fields over the contiguous United States

Jiali Wang[1] and Veerabhadra R. Kotamarthi[1]

[1]Environmental Science Division, Argonne National Laboratory, Argonne, Illinois, USA

Abstract The Weather Research and Forecasting (WRF) model is used for dynamic downscaling of 2.5-degree National Centers for Environmental Prediction-U.S. Department of Energy Reanalysis II (NCEP-R2) data for 1980–2010 at 12 km resolution over most of North America. The model's performance for surface air temperature and precipitation is evaluated by comparison with high-resolution observational data sets. The model's ability to add value is investigated by comparison with NCEP-R2 data and a 50 km regional climate simulation. The causes for major model bias are studied through additional sensitivity experiments with various model setup/integration approaches and physics representations. The WRF captures the main features of the spatial patterns and annual cycles of air temperature and precipitation over most of the contiguous United States. However, simulated air temperatures over the south central region and precipitation over the Great Plains and the Southwest have significant biases. Allowing longer spin-up time, reducing the nudging strength, or replacing the WRF Single-Moment six-class microphysics with Morrison microphysics reduces the bias over some subregions. However, replacing the Grell-Devenyi cumulus parameterization with Kain-Fritsch shows no improvement. The 12 km simulation does add value above the NCEP-R2 data and the 50 km simulation over mountainous and coastal zones.

1. Introduction

As interest in and concern about regional climate change increase, the desire for greater regional detail in climate model projections is expected to grow. Assessment of climate impacts and development of adequate adaptation measures are among the greatest and most important challenges facing mankind in the coming decades [*Mearns et al*., 2012; *Salzmann and Mearns*, 2012]. However, climate simulations by conventional general circulation models (GCMs) with relatively coarse spatial resolution are not necessarily capable of capturing the details associated with regional-local climate variability and the changes required for regional and national climate change assessments [*Giorgi and Mearns*, 1999; *Denis et al*., 2002; *Giorgi et al*., 2009; *Rummukainen*, 2010]. In addition, at coarse grid resolutions, the magnitude and intensity of subgrid-scale extreme events such as heavy rainfall (leading to floods) are often not captured nor realistically reproduced [*Tripathi and Dominguez*, 2013].

Regional climate models (RCMs) with considerably higher resolution are therefore constructed for limited areas to describe regional-scale climate variability and change. The RCMs are constrained at the boundaries and partially over the inner domain by global- or relatively large-scale driving data. This technique, called dynamical downscaling, has proven to be a useful tool for generating climate projections at higher spatial resolution by downscaling reanalysis data sets [e.g., *Castro et al*., 2005; *Martynov et al*., 2013]. Several dynamical downscaling studies have been conducted over the North American domain for historical and/or future climate [e.g., *Mearns et al*., 2007, 2009; *Wi et al*., 2012; *Martynov et al*., 2013]. Among these, the North American Regional Climate Change Program (NARCCAP) (www.narccap.ucar.edu) is the most widely available dynamical downscaling product. The NARCCAP has employed six RCMs driven by National Centers for Environmental Prediction-U.S. Department of Energy Reanalysis II (NCEP-R2) data for a 25 year historical period (1980–2004) and for selected future climate scenarios with the Intergovernmental Panel on Climate Change A2 emission scenarios. *Mearns et al*. [2012] evaluated the performance of these RCMs for annual temperature and precipitation cycles over the entire domain and four subdomains—southern California, Great Plains, south central, and Atlantic coast—and found that the seasonal temperature was relatively well reproduced by most models, but seasonal precipitation was less successfully captured. For example, the

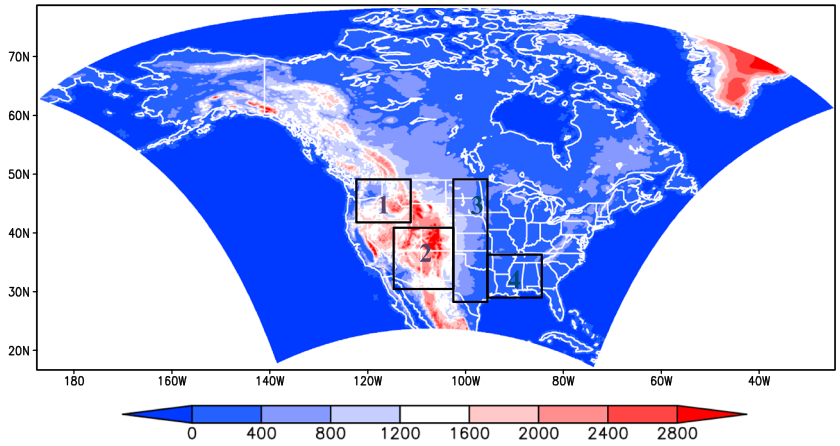

Figure 1. The WRF model domain with terrain height elevations (m) indicated by shading, for four subregions: 1, North Rockies (NR); 2, Southwest (SW); 3, Great Plains (GP); and 4, South Central (SC).

Weather Research and Forecasting (WRF) model's result had ~40% dry bias over the eastern contiguous United States (CONUS) in the summer [*Mearns et al*., 2012, Figure 10f], as well as 20–40% dry bias over the south central region and >60% wet bias over western CONUS in the winter [*Mearns et al*., 2012, Figure 9f]. *Martynov et al*. [2013] generated and evaluated a 50 year simulation (1959–2008) over the Coordinated Regional Climate Downscaling Experiment North American domain by using the Canadian Regional Climate Model (version 5) and found that the model had a cold bias (2–6°C) over the Rocky Mountains in the summer and year-round over Mexico, plus a wet bias (50–75%) over the Great Plains and mountainous and coastal regions in the winter. In addition to the bias due to physics parameterizations applied in these RCMs, the biases over mountainous regions could be due to the coarse grid spacing. In these regions the model cannot resolve subgrid scale topographic features and hence model the additional surface drag, introducing errors into the atmospheric circulations. On the other hand, the biases over the south central region could result from drifts in atmospheric variables (e.g., wind, temperature, and pressure) caused by errors accumulated during the continuous integration.

This study uses the WRF model for simulating the regional climate over most of North America. The model domain (Figure 1) is described in section 2. Our work differs from prior studies in several key aspects. First, the model grid spacing is significantly finer over such a large domain, at a spatial resolution of 12 km, versus 50 km used for the NARCCAP, 35 km used by *Wi et al*. [2012], and ~48.4 km used by *Martynov et al*. [2013]. Second, application of the spectral nudging technique results in significant improvements in model performance for near-surface precipitation, air temperature, and wind [*Wang and Kotamarthi*, 2013]. In contrast, some NARCCAP RCMs and the simulation of *Martynov et al*. [2013] did not apply spectral nudging, which could cause model drift due to accumulated errors during long integration times. Third, we present an in-depth analysis of the dominant reasons for the biases in our calculations, conducted by setting up additional sensitivity experiments and investigating the related physical processes. This is important for allowing the dynamical downscaling community to understand model performance and to set up an optimal model simulation.

We wish to accomplish three goals: (1) demonstrate that the 12 km WRF simulation captures the essential features of spatial patterns and annual cycles in air temperature and precipitation, on the basis of quantitative statistical approaches; (2) demonstrate that the 12 km WRF simulation adds value above the driving data and coarse-resolution simulation by comparison with NCEP-R2 data and the NARCCAP-WRFG (WRF Grell [*Grell and Devenyi*, 2002]) RCM (www.narccap.ucar.edu) in historical periods; and (3) explore the dominant reasons for the model biases by designing additional sensitivity experiments and then suggest an approach for constructing the WRF model to achieve optimal performance in near-surface fields.

Section 2 describes the model setup, experimental designs, and evaluation data sets. Sections 3 and 4 evaluate the model performance quantitatively for temperature and precipitation, respectively. The potential reasons for the model biases and related physical processes are explored and discussed in section 5. A summary and further discussions about possible reasons for the WRF improvements over the NARCCAP-WRFG are in section 6.

2. Model Description, Experimental Design, and Evaluation Data Sets

2.1. Model Description

In this work, we use the WRF with the Advanced Research WRF dynamic core, version 3.3.1. The simulation domain is centered at 52.24°N and 105.5°W and has dimensions of 600×516 horizontal grid points in the west-east and south-north directions with grid spacing of 12 km, covering most of North America (Figure 1). The calculations and results from this model are referred to as "the WRF model" in the rest of the paper. The initial and boundary conditions are constructed from NCEP-R2 data [*Kanamitsu et al*., 2002]. The model is reinitialized (for both atmospheric and surface forcing) every year during 1980–2010. Such a sequence of relatively short runs with numerous reinitializations was shown by *Pan et al*. [1999], *Qian et al*. [2003], and *Lo et al*. [2008] to outperform the traditional long-term continuous simulations for not only the forcing variables (e.g., pressure, temperature, wind, and moisture) but also the model diagnostic variables (e.g., precipitation). The approach is becoming increasingly accepted and adopted [e.g., *Conil and Hall*, 2006; *Zagar et al*., 2006; *Jiménez et al*., 2010; *Lucas-Picher et al*., 2013]. However, the model needs to adjust (spin-up) for each reinitialization. To minimize the adjustment problems, *Pan et al*. [1999] allowed a 3 day spin-up time for each 13 day reinitialization in one of their experiments. They found that this simulation performed similarly to a 30 day continuous run, which was less successful than the 10 day reinitialization without spin-up time.

Qian et al. [2003] detected two stages of temporal adjustment in the model spin-up process. Rapid adjustment in the first day or two to a state intrinsic to the model dynamics resulted from removal of small-scale imbalances through geostrophic adjustments from mass to wind field. In the subsequent 15–20 days, the model variables adjusted gradually to a state close to that of the long-term continuous run. After this, the simulation results were almost identical to those for the continuous run, which showed larger bias than the reinitialization. Therefore, in this study, to reduce the adjustment problems introduced by each reinitialization and to maximize the strengths of reinitialization, we allow 24 h for each reinitialization, in agreement with the spin-up time used by *Lo et al*. [2008] in their reinitializations. Moreover, *de Ela et al*. [2002] and *Lucas-Picher et al*. [2013] showed that most of the small-scale variabilities and vorticities were fully generated in this time. Another advantage of reinitialization is the ability to run the model in parallel over different periods. On the other hand, reinitialization of the surface forcing is reasonable, because *Pan et al*. [1999] found that soil moisture was less important than atmospheric forcing by comparing two experiments—one reinitializing all the forcing fields and the other only reinitializing the atmospheric variables and keeping the soil moisture uninterrupted.

For each year during 1980–2010, the model runs are continuous, with spectral nudging but without reinitialization, because *Lo et al*. [2008] found that it was not necessary to subdivide integrations into relatively short runs when interior nudging was applied. The variables spectrally nudged in this study are horizontal winds, temperature, and geopotential height. Spectral nudging is applied above 850 hPa to wavelengths around 1200 km with a nudging coefficient of $3 \times 10^{-4}\,s^{-1}$. The minimum wavelength for spectral nudging corresponds to the minimum wavelength resolved in the input fields, and the minimum wavelength resolved should be at least $4\Delta x$ [*Pielke*, 2002; *Castro et al*., 2005; *Rockel et al*., 2008], which is ~1100 km for NCEP-R2 in midlatitudes. The selection of nudging wavelength and the nudging height for generating the most realistic surface temperature, winds, and precipitation was described by *Wang and Kotamarthi* [2013].

The physical options incorporated into this study are the WRF Single-Moment six-class graupel scheme (WSM6) [*Hong and Lim*, 2006], the Grell-Devenyi cumulus parameterization (G-D) [*Grell and Devenyi*, 2002], the Rapid Radiative Transfer Model for GCMs longwave and shortwave schemes (RRTMG, http://rtweb.aer.com) [*Iacono et al*., 2008], the Yonsei University planetary boundary layer (PBL) scheme [*Noh et al*., 2003], and a land surface model by *Chen and Dudhia* [2001].

2.2. Experimental Design

To explore the dominant reasons for the WRF model bias, we design four additional sensitivity experiments with various physics parameterizations and model setup approaches in an arbitrarily chosen year—2005—that shows typical biases of the long-term means (section 5), as follows:

1. In Test 1, we replace the G-D cumulus parameterization with the Kain-Fritsch (K-F) [*Kain and Fritsch*, 1990, 1993; *Kain*, 2004] to investigate the model's sensitivity to cumulus parameterizations in simulating summer precipitation, because most of the summer precipitation over the United States is induced by mesoscale or small-scale convective systems [*Liang et al*., 2004a; *Castro et al*., 2007; *Tripathi and Dominguez*, 2013].

2. In Test 2, we reduce the nudging coefficient from $3\times10^{-4}\,s^{-1}$ to $3\times10^{-5}\,s^{-1}$, to test the hypothesis that very strong nudging destroys the mesoscale features generated by the fine-scale models [*Bowden et al.*, 2012; *Otte et al.*, 2012; *Omrani et al.*, 2012; *Tripathi and Dominguez*, 2013].
3. In Test 3, we replace the WSM6 microphysics scheme with that of *Morrison et al.* [2009] to investigate the model's sensitivity to the microphysics scheme in simulating winter precipitation. *Liu et al.* [2011] found that the WRF Single-Moment schemes tend to overpredict winter precipitation over mountain ranges of the western United States, and the Morrison scheme was shown to address this problem. Other physics parameterizations (e.g., land surface model, PBL schemes, and radiation schemes) have only moderate to weak impacts on the modeled precipitation over this region [*Liu et al.*, 2011].
4. In Test 4, we run a 2 year continuous simulation (2004–2005) with 1 year (2004) spin-up time and compare the other year (2005) with the reinitialized 2005. This is because *Pan et al.* [1999] found that reinitialization tended to produce excessive rainfall, although the rainfall position is improved. On the other hand, soil parameters are "remembered" longer in models than are the atmospheric variables. Thus, soil parameters probably need longer spin-up times [*Lucas-Picher et al.*, 2013; *Tripathi and Dominguez*, 2013], although they are secondary in importance to atmospheric forcing for model performance [*Pan et al.*, 1999].

2.3. Evaluation Data Sets

Monthly temperature and precipitation values from the Precipitation-Elevation Regressions on Independent Slopes Model (PRISM) developed by *Daly et al.* [1994, 1997, 2008] are used to evaluate the model's performance in the annual cycle. The PRISM values, which are corrected for systematic elevation effects on precipitation climatology, provide observation-based temperature/precipitation on a grid mesh of 1/8° latitude × 1/8° longitude that covers all of CONUS. The PRISM monthly average temperature is calculated from the maximum and minimum temperatures (Tmax and Tmin, average = (Tmax + Tmin)/2), as recommend by the PRISM group (http://www.prism.oregonstate.edu/faq.phtml). To explore potential differences between the true monthly average temperature and the average of Tmax and Tmin, we compare the two values over each station (with thousands of station-observed daily values from the National Climate Data Center) in the U.S. Midwest and Southwest by day in 1980–2004. The results show that more than 99% of the differences between the two values are ~0.05°C, which we believe would not influence the evaluation of the model's performance for surface air temperature. Given the strong dependence of temperature and precipitation on elevation, the topographic adjustment is critical, because cooperative stations over mountainous regions are preferentially located at lower elevations and thus tend to underestimate (overestimate) the true spatial average of precipitation (temperature). Therefore, the PRISM observation-based values are the most accurate for estimating precipitation and temperature, especially in topographically complex regions. PRISM is also the best available data set for evaluating high-resolution model simulations. Other widely used data sets, such as the University of Delaware values [*Willmott and Matsuura*, 1995; *Matsuura and Willmott*, 2010] and the Climatic Research Unit data [*Mitchell and Jones*, 2005; *Harris et al.*, 2013], are not necessarily able to resolve detailed information as highly as the WRF model used in this study. The NCEP North American Regional Reanalysis (NARR) [*Mesinger et al.*, 2006; *Bukovsky and Karoly*, 2007] 3 h precipitation (assimilated by hourly observed precipitation)—the best available gridded data set at the diurnal scale—is used to evaluate the diurnal cycles in precipitation.

Because the primary evaluation data sets are for regions within CONUS, the model evaluation region used in this study is also confined to CONUS. Four subregions with various model bias and climatic and topographic features are selected for analysis. These subregions are identified as the Great Plains (GP), Southwest (SW), South Central (SC), and North Rockies (NR) subregions (Figure 1).

To retain the original features of the data sets at various resolutions during model evaluation and comparison with other data sets, we use the native resolution (with no interpolation) of different data sets/models to plot geographic patterns and calculate statistical metrics except for spatial/temporal correlation coefficients and precipitation differences between model simulations and PRISM, which are computed by regridding the PRISM data from 4 km to the WRF at 12 km and to the NARCCAP-WRFG at 50 km.

3. Evaluation and Comparison of Temperature

In section 3, the model's capability to reproduce the annual cycle in air temperature is assessed through comparison with the PRISM observations, the NCEP-R2 reanalysis, and the NARCCAP-WRFG simulation. The

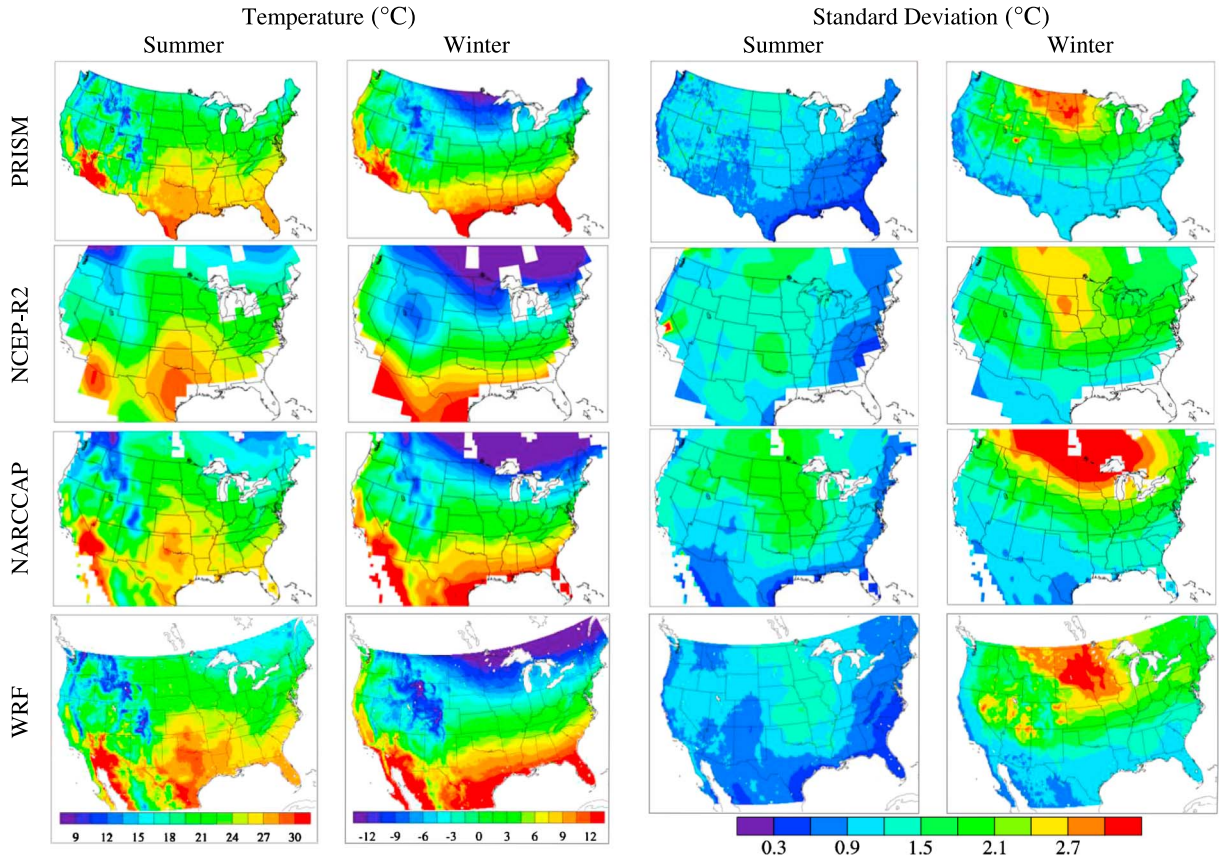

Figure 2. Multiannual (1980–2004) (first and second columns) average and (third and fourth columns) standard deviation of summer and winter 2 m temperatures (°C) from (first row) PRISM, (second row) NCEP-R2, (third row) NARCCAP-WRFG, and (fourth row) WRF.

NARCCAP-WRFG used the same convective parameterization (G-D) as the WRF model and was also driven by NCEP-R2 data. However, the studies differ in domain size (see Figure 1 in this study and Figure 1 in *Mearns et al*. [2012]), nudging technique (turned on in this study and off in the NARCCAP-WRFG), shortwave and longwave radiation schemes (RRTMG in this study but National Center for Atmospheric Research (NCAR) Community Atmospheric Model radiation scheme in the NARCCAP-WRFG), and approach to integrating the model (reinitialization of the model in this study versus continuous running the model in the NARCCAP-WRFG). These differences make a direct comparison of simulated climatology difficult. In sections 3 and 4, we compare the performance of the WRF and NARCCAP-WRFG to show potential improvements/shortcomings. However, it should be noted that the differences between the results are not directly attributable to a particular difference between the models but are a combined effect of all the differences. To make this comparison structured and illustrative, in section 6.2, we investigate the effect of each of these differences between the models and identify the likely combination of changes that leads to the net calculated difference between the models. We select the overlap years (1980–2004) between the WRF and the NARCCAP-WRFG for comparison.

Figure 2 (first and second columns) compares multiannually averaged 2 m temperature from PRISM, WRF, NARCCAP-WRFG, and NCEP-R2 in summer (June, July, and August, JJA) and winter (December, January, and February, DJF). Because of its coarse resolution (2.5° × 2.5°), NCEP-R2 cannot resolve mesoscale orographic temperature patterns over mountain ranges as well as over coastal zones, and thus it shows a single broad center over all major mountain ranges and significant differences in temperature during all the seasons over the Rockies. It also underestimates temperature in all four seasons (spring and fall not illustrated) over southern California and Arizona. The NARCCAP-WRFG can resolve part of the orographic temperature pattern over the mountain ranges and coastal areas, showing more details of temperature than the NCEP-R2. Both NCEP-R2 and NARCCAP-WRFG show bias (±1.5°C) over Texas and Florida in

summer. Though the WRF model underestimates the temperature over the western mountain ranges by 1.5–3.0°C in both seasons, it captures many more details of the temperature patterns than do the NCEP-R2 and NARCCAP-WRFG, especially over the mountain ranges and the coastal zones, showing topographic features similar to those in the PRISM observations. On the other hand, the NARCCAP-WRFG overestimates the temperature over the northern GP and the Midwest in spring (not shown) and in winter by ~3°C and underestimates the temperature over Texas, Louisiana, and Florida in summer by ~2°C. The WRF shows much smaller bias (~1°C) and closer results to the PRISM values than does the NARCCAP-WRFG. The spatial correlations between the PRISM values and the WRF results are 0.997 and 0.988, while those between the PRISM values and the NARCCAP-WRFG results are 0.945 and 0.975 in summer and winter, respectively (Table S1 in the supporting information).

Figures 3a–3d compare the subregional-average bias of temperature for WRF, NARCCAP-WRFG, and NCEP-R2 versus PRISM observations in all four seasons during 1980–2004. The error bars denote the yearly distribution of the bias at the 10th and 90th percentiles. Generally, the annual variations in the biases for WRF and NCEP-R2 are smaller than those for NARCCAP-WRFG. Improvements of WRF over NARCCAP-WRFG are seen over GP (Figure 3a) in spring and winter, over SW (Figure 3b) in all four seasons, and over NR (Figure 3d) in spring, summer, and winter. For example, over SW, the NARCCAP-WRFG bias is 2.8–4.4°C, while the WRF bias is only 0.06–1.0°C; over NR, the NARCCAP-WRFG bias is 0.3–2.6°C, while the WRF bias is 0.2–2.0°C (except in fall). Topography plays a key role in determining temperature over mountain ranges, and (as expected) the WRF model with its better representation of topography leads to smaller temperature root-mean-square errors (RMSEs) than the NARCCAP-WRFG over SW and NR (Figure S1 in the supporting information). For example, over SW, the RMSEs for NARCCAP-WRFG are 2.9–4.5°C, while those for WRF are only 0.3–1.2°C; over NR, the RMSEs for NARCCAP-WRFG are 0.8–2.8°C, while those for WRF are 0.4–2.1°C. However, WRF shows larger warm biases than NARCCAP-WRFG over GP (Figure 3a) in summer and fall and over SC (Figure 3c) in all four seasons. The RMSEs for WRF over SC are also larger than those for NARCCAP-WRFG in all seasons except summer (Figure S1). The possible reasons for the WRF model biases are explored in section 5, and the main reasons for the WRF improvements over NARCCAP-WRFG are discussed in section 6.2.

To achieve good fidelity for the WRF model in simulating climate, we expect the model not only to capture the mean fields and reduce the model bias, but also to generate variances similar to those of the observations. Thus, in addition to assessing mean fields, we compare the standard deviation (SD) of temperature for WRF, NARCCAP-WRFG, and NCEP-R2 with the SD of the PRISM observations, as shown in Figure 2 (third and fourth columns). In summer, the NCEP-R2 overestimates the SDs over most of CONUS by 0.3–0.6°C, and it does not capture the fine-scale features over mountains and coasts. The NARCCAP-WRFG shows a closer SD pattern to the PRISM but overestimates the SDs by 0.6–0.9°C. In winter, the NCEP-R2 overestimates the SDs over southern CONUS by ~0.6°C, while it underestimates the SDs over South Dakota, North Dakota, and Montana by 0.6°C. The NARCCAP-WRFG shows a similar SD pattern to the PRISM but overestimates the SDs over northern CONUS by 0.6–0.9°C. The WRF captures the spatial pattern and the amplitude of temperature SDs very well in all four seasons (spring and fall not illustrated), with the smallest bias and the closest pattern to the PRISM in comparison with the NCEP-R2 and the NARCCAP-WRFG. The improvement of the WRF performance in SDs over NARCCAP-WRFG probably results from the spectral nudging [*Feser*, 2006; *Omrani et al*., 2013] and the reinitialization applied in the WRF.

4. Evaluation and Comparison of Precipitation

4.1. Seasonal Pattern

Figure 4 (first and second columns) compares multiannual average precipitation from PRISM, WRF, NARCCAP-WRFG, and NCEP-R2 in summer and winter. The NCEP-R2 with coarse resolution does not resolve the mesoscale orographic precipitation patterns over western CONUS with complex topography. Precipitation is underestimated over both the Cascade Range and the Rocky Mountains. The NARCCAP-WRFG captures more details of the precipitation pattern over these regions than does NCEP-R2. The WRF generates more details than NCEP-R2 and NARCCAP-WRFG, and the results are closer to the PRISM values over the mountainous regions and coast zones in all four seasons. Spatial correlation coefficients (Table S1) between the WRF results and PRISM observations are 0.946 in summer and 0.963 in winter, while those between the NARCCAP-WRFG results and PRISM observations are 0.725 in summer and 0.812 in winter.

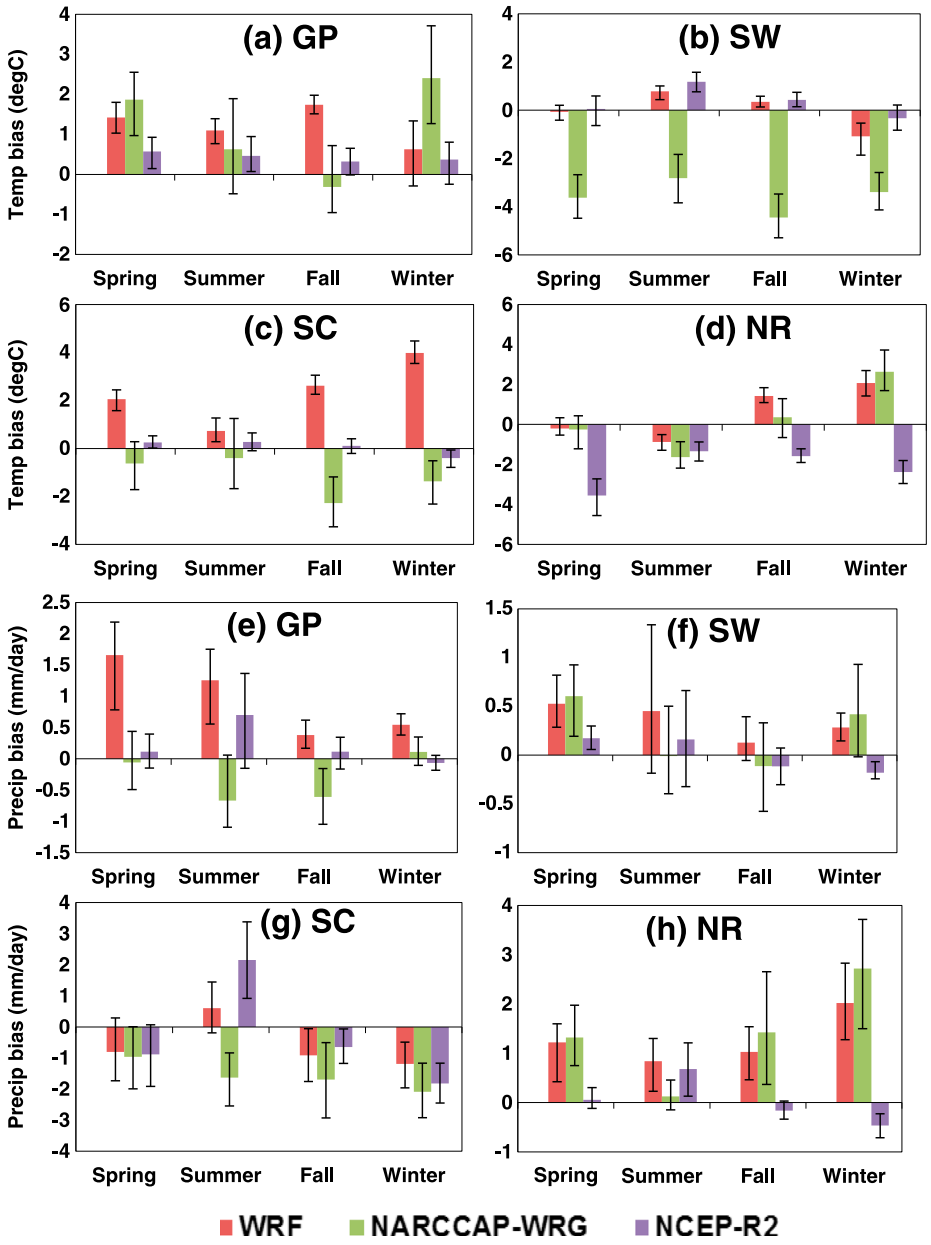

Figure 3. Subregional-average bias of (a–d) 2 m temperature and (e–h) precipitation from WRF, NARCCAP-WRFG, and NCEP-R2 versus PRISM data by season in 1980–2004 over the Great Plains (Figures 3a and 3e), Southwest (Figures 3b and 3f), South Central (Figures 3c and 3g), and North Rockies (Figures 3d and 3h) subregions. The error bars denote the annual distribution of bias at the 10th and 90th percentiles.

Figures 3e–3h show the subregional-average precipitation bias and its yearly distribution in 1980–2004 over the four subregions. The corresponding percent bias values, defined as in *Mearns et al.* [2012], are shown in Table S2. Both the WRF and NCEP-R2 show a wet (dry) bias in summer (winter) over SC (Figure 3g) that has also been reported in GCMs [*Liang et al.*, 2004a]. The NARCCAP-WRFG shows dry bias in all four seasons over SC. An encouraging improvement is that WRF shows smaller biases (Figure 3g) and RMSEs (Figure S1) than NCEP-R2 in all seasons except summer (and also fall for RMSEs). In addition, WRF shows significantly smaller biases and RMSEs than NARCCAP-WRFG in all four seasons. As Figure 3g shows, in spring, summer, fall, and winter, respectively, the bias of the NCEP-R2 over SC is −0.9 (−21.6%), 2.2 (62.5%), −0.6 (−17.5%), and −1.8 (−45.1%) mm d^{-1}; the NARCCAP-WRFG bias over SC is −1.0 (−23.4%), −1.6 (−47.2%), −1.7 (−46.0%), and −2.1 (−51.7%) mm d^{-1}; and the WRF bias over SC is −0.8 (−19.6%), 0.6

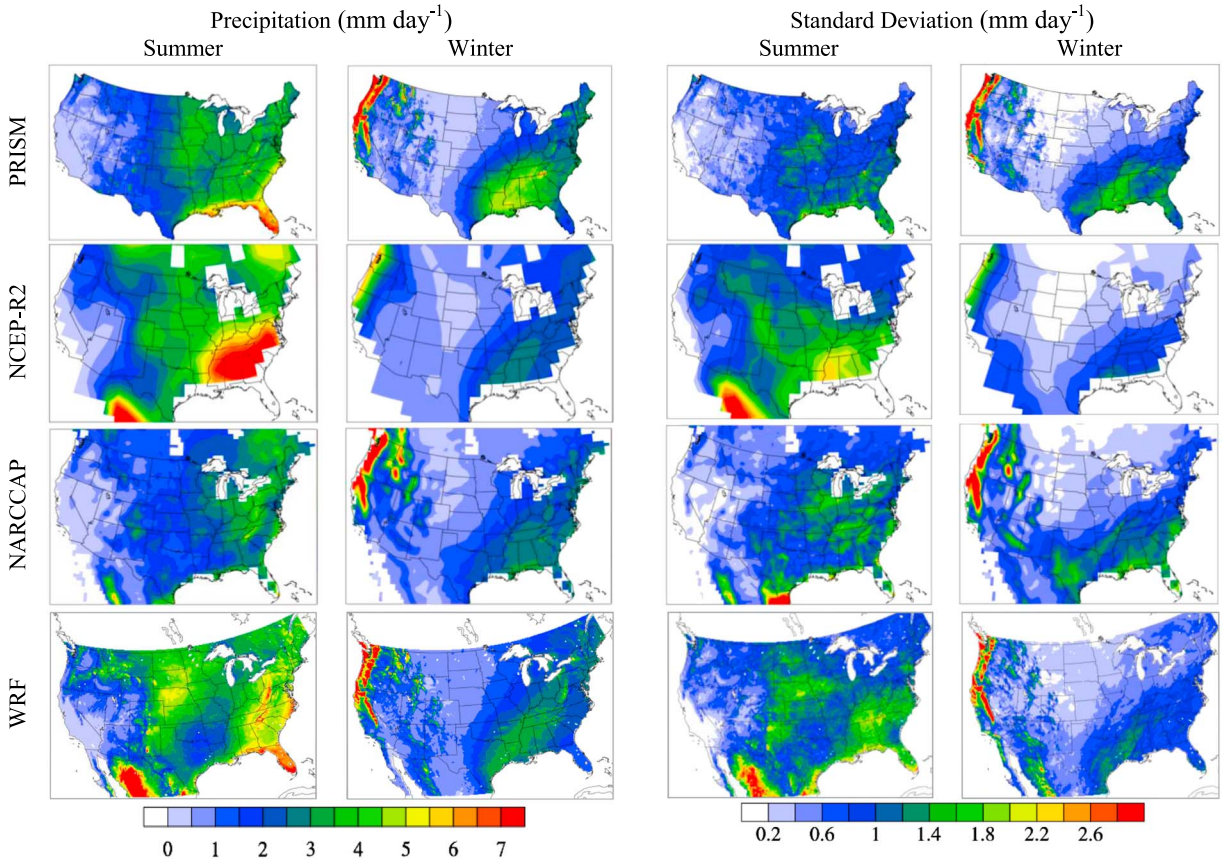

Figure 4. Multiannual (1980–2004) (first and second columns) average summer and winter precipitation and (third and fourth columns) standard deviation (mm day^{-1}) from (first row) PRISM, (second row) NCEP-R2, (third row) NARCCAP-WRFG, and (fourth row) WRF.

(17.5%), −0.9 (−24.8%), and −1.1 (−29.5%) mm d^{-1}. On the other hand, WRF generates smaller wet bias than NARCCAP-WRFG over NR (Figure 3h)—1.1 (70.4%), 1.0 (72.2%), and 2.0 (108%) mm d^{-1} versus 1.3 (76.3%), 1.5 (100%), and 2.7 (146%) mm d^{-1} in spring, fall, and winter, respectively, when the dominant rainfall occurs over this region. The WRF bias over SW (Figure 3f) is also smaller than the NARCCAP-WRFG bias—0.5 (54.7%) and 0.3 (32.2%) mm d^{-1} versus 0.6 (62.7%) and 0.4 (47.3%) mm d^{-1} in spring and winter, respectively.

The RMSEs (Figure S1) over SW and NR are similarly improved by WRF over NARCCAP-WRFG. We hypothesize that the improvement results from the reinitialized WRF at high resolution, with spectral nudging being better than NARCCAP-WRFG at capturing atmospheric circulations over the mountain ranges and generating precipitation patterns closer to the observations, as discussed further in section 6.2. However, WRF overestimates rainfall over GP in all seasons (Figures 3e and 4, first and second columns) and shows improvement over NARCCAP-WRFG only in fall, decreasing the absolute bias from 0.6 (41.5%) to 0.4 (25.9%) mm d^{-1}. The possible reasons for this significant wet bias are investigated by designing additional sensitivity experiments (section 5).

As for temperature, we also compare the SDs of precipitation between PRISM, WRF, NARCCAP-WRFG, and NCEP-R2 (Figure 4, third and fourth columns). In summer, the NCEP-R2 overestimates the SDs by >100% over most of CONUS, especially over northern GP and southeastern CONUS. The WRF shows overestimations over similar regions to the NCEP-R2, while NARCCAP-WRFG shows overestimations over the Midwest and northeastern CONUS. In winter, the largest SDs distribute over Cascade Range and the Rocky Mountains. Both WRF and NARCCAP-WRFG capture this, while NCEP-R2 shows a 30% underestimation. WRF adds many more details over the mountain ranges and coastal zones and is closer to the PRISM values than NCEP-R2 and NARCCAP-WRFG. However, both NCEP-R2 and WRF underestimate the SDs by ~50% over southeastern CONUS, while NARCCAP-WRFG overestimates the SDs over Texas by ~50%. The bias of precipitation SDs

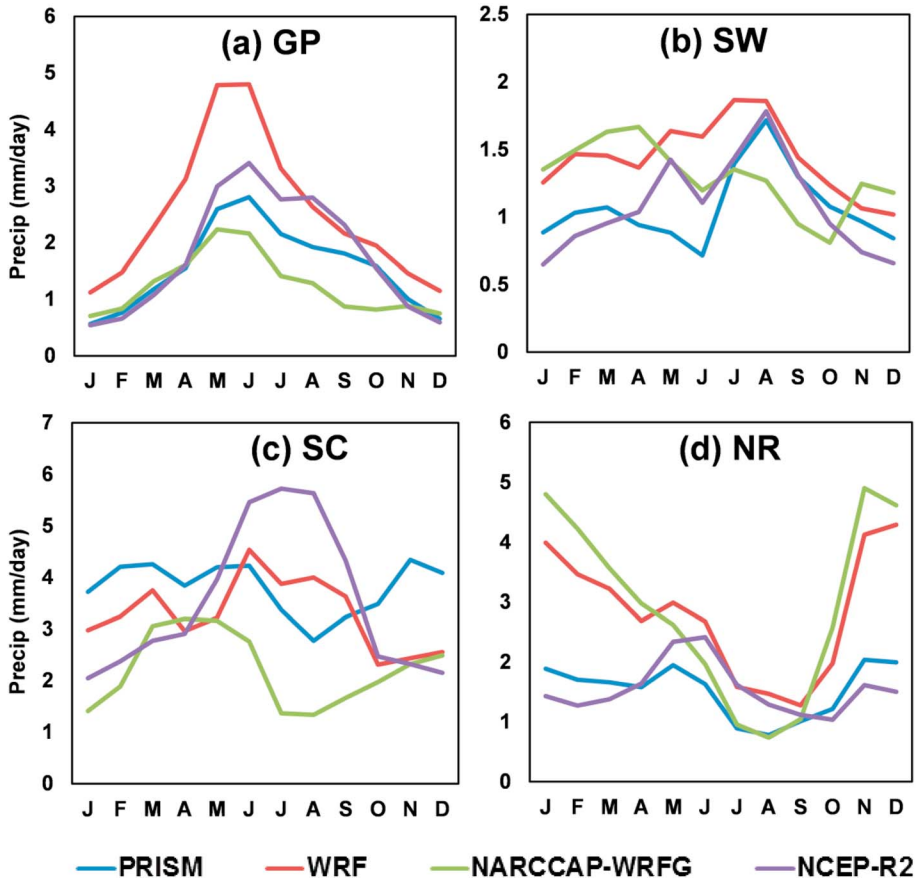

Figure 5. Multiannual (1980–2004) average monthly variations in precipitation over the (a) Great Plains, (b) Southwest, (c) South Central, and (d) North Rockies subregions from PRISM, WRF, NARCCAP-WRFG, and NCEP-R2.

probably has the same causes as the bias in seasonal averages of precipitation, as the bias of SDs occurs over the same regions as the bias of averages.

4.2. Monthly Variation

In addition to the climatological seasonal pattern, evaluation of monthly variation is important, because seasonal averages could hide biases that are identifiable in monthly data. Figure 5 compares the monthly variations of precipitation for PRISM, NCEP-R2, NARCCAP-WRFG, and WRF over the four subregions specified in Figure 1. Like the PRISM observations, WRF generates a rainfall peak in May–June over GP and in July–August over SW. This similarity in the monthly transition in precipitation is encouraging, because it is an important indicator of the pre-North American Monsoon (NAM) and the onset and peak of NAM [*Castro et al.*, 2007]. Further detailed evaluation [e.g., *Lebassi-Habtezion and Diffenbaugh*, 2013] of the model fidelity in the NAM will be conducted in our future work. NARCCAP-WRFG does not capture the NAM rainfall transition well, with a rainfall peak in April and minimum rainfall in October over SW. However, over GP (Figure 5a) WRF produces much heavier rainfall than the PRISM values from January to July. The possible reasons for this wet bias are investigated and discussed in sections 5 and 6. Over SC (Figure 5c), the NCEP-R2 generates substantial wet (dry) bias in the summer (winter) month—opposite the PRISM values. NARCCAP-WRFG presents a monthly variation similar to that of PRISM but shows a significant dry bias during the entire year. Especially in January–March, WRF produces heavier rainfall than NARCCAP-WRFG, closer to the PRISM values. However, WRF generates less (more) rainfall in the cold (warm) season than the PRISM values. Over NR (Figure 5d), both NARCCAP-WRFG and WRF show wet biases in the cold season, when the region receives most of its annual precipitation. WRF shows slightly better monthly variation than NARCCAP-WRFG during the cold season, with less rainfall (closer to the PRISM values). The possible reasons for this wet bias during the winter over NR are investigated in section 5.

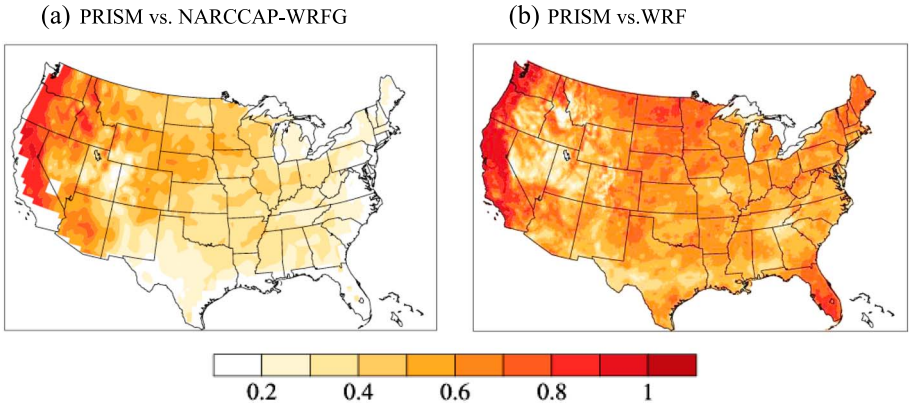

Figure 6. Temporal correlation coefficients in precipitation (a) between PRISM and NARCCAP-WRFG at 50 km and (b) between PRISM and WRF at 12 km during all months in 1980–2004. The Student's *t* test result at the 0.005 (and lower) level of significance is marked by the color scale.

The temporal correlation coefficients (TCCs) between simulated and observed monthly precipitation over each grid are significantly improved by WRF in comparison with NARCCAP-WRFG (Figure 6). WRF captures the monthly variations in precipitation over CONUS, with TCC $>$ 0.5 (significant level = 0.005) over most of CONUS, except over the Rockies. NARCCAP-WRFG shows significant TCC values over western CONUS, but the lower TCC values over eastern CONUS, especially close to the ocean and the Gulf of Mexico, do not pass the statistical significance test.

5. Possible Reasons for the Model Bias

The diurnal cycle of precipitation, in comparison with the subsynoptic and synoptic cycles, is the dominant mechanism for summer precipitation over most of CONUS, as shown by *Castro et al.* [2007] on the basis of spectral analysis of integrated moisture flux convergence. Our model simulation also indicates the dominance of convective precipitation (>60% of total precipitation) in summer. From rain gauge observations and satellite detections, *Liang et al.* [2004b] and *Tian et al.* [2005] found that the rainfall typically peaked over the GP around midnight, over the southeastern United States from roughly sunset and into the night, and over the Midwest in the night to early morning. *Tripathi and Dominguez* [2013] found that their model produced slower propagation of the rainfall system and heavier hourly rainfall than observations over the U.S. Southwest, thus inducing a wet bias in the climatological daily precipitation.

On the other hand, several aspects of model setup (e.g., spectral nudging and reinitialization) and physics parameterizations could affect the pattern and amplitude of climate variables. In this study, the most significant problem with the model is the wet bias over GP in the warm season and the warm bias over SC in all four seasons. Also, there is still a wet bias over the mountain ranges in cold seasons although the WRF shows pronounced improvement over the NARCCAP-WRFG results (Figures 3f, 3h, 5b, 5d, and S1). Therefore, to explore the potential reasons for the model biases, we first compare the diurnal variations in precipitation between the NARR values and the WRF results over CONUS. Then we conduct several experiments to test the model sensitivities to convective parameterization, microphysics scheme, spectral nudging strength, and spin-up time.

5.1. Diurnal Variation of Summer Precipitation

Figure 7 shows the climatological diurnal cycles of precipitation in May–August over CONUS from the NARR values (first and third rows) and the WRF simulation (second and fourth rows). Since the NARR assimilates hourly precipitation, it captures the diurnal cycle of precipitation very well in comparison with observations, as shown by *Liang et al.* [2004b] and *Tian et al.* [2005]. The rainfall peak over GP occurs at around 0600–0900 UTC, and then rainfall systems move eastward and generate a rainfall peak over the Midwest at around 1200 UTC. Later, rainfall reaches a maximum over southeastern CONUS at 1800–2100 UTC. The WRF precipitation propagation eastward is much slower than in the NARR values. For example, the rainfall band has almost left the GP and gone into the Midwest by 1200 UTC in NARR, whereas for WRF it persists until 1500 UTC.

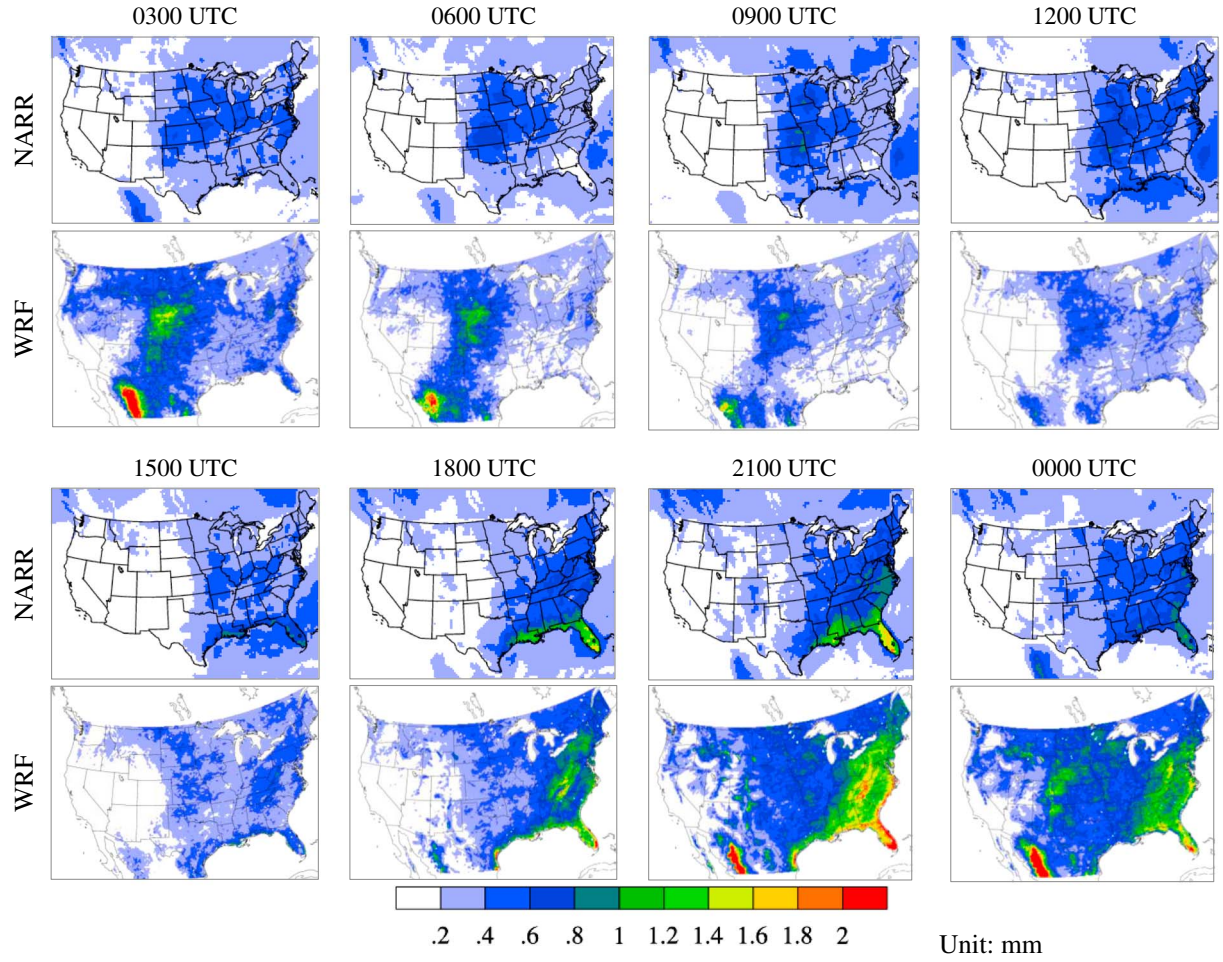

Figure 7. Climatological diurnal cycles of (second and fourth rows) WRF-simulated precipitation and (first and third row) NARR values for May–August.

In addition to the problem of propagation speed, the rainfall center over GP is more northward (over Nebraska) than in the NARR values (over Kansas). Strong wet biases over the GP and over northwestern CONUS are evident in the core of the rainfall system at all evaluation times. The WRF captures the diurnal precipitation cycle over southeastern CONUS but overestimates the rainfall amounts. Both the slow propagation of the rainfall system and the overestimation of precipitation induce the wet bias for daily mean precipitation over GP and western CONUS, and they likely result from the model setup and/or the physics parameterizations applied in our simulation.

5.2. K-F Convective Parameterization

To explore the possible reasons for the model biases discussed above, a sensitivity experiment (Test 1) replacing the G-D cumulus parameterization with the K-F parameterization is conducted for summer 2005 (JJA). The precipitation bias in summer 2005 is typical of the long-term summer means (Figure 8a), with wet bias over the northern GP, the southern U.S. mountains, SW, and NR, as well as dry bias over Kansas, Oklahoma, Missouri, and Arkansas. Figure 8b illustrates the rainfall difference between Test 1 (K-F cumulus parameterization) and the PRISM observation. Use of K-F instead of G-D does not result in significant improvement, but even stronger wet biases are calculated over the Northwest, SW, and Southeast.

The strong wet biases generated by Test 1 are mostly (>90%) contributed by convective precipitation. Namely, Test 1 with K-F scheme produces more convective precipitation while producing slightly less microphysics-resolved precipitation. One reason for the increased convective precipitation is that Test 1 (K-F) generates higher convective available potential energy (CAPE) than the control simulation (G-D) over SC. This was also found by *Liang et al.* [2004a] when they replaced the G-D scheme with K-F in their model simulation.

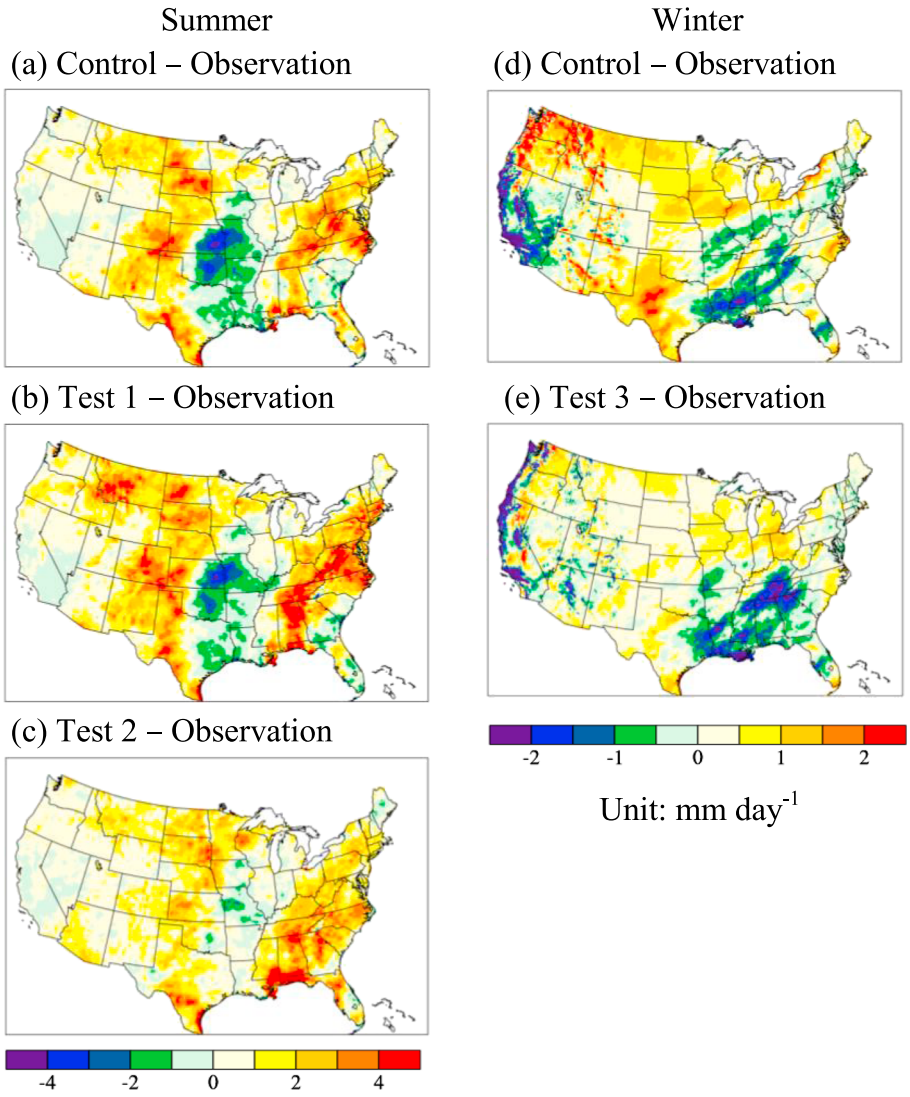

Figure 8. Precipitation in summer (Figures 8a–8c) and winter (Figures 8d and 8e) 2005 for the (a, d) difference between the control simulation and PRISM observation, (b) difference between Test 1 and PRISM observation, (c) difference between Test 2 and PRISM observation, and (e) difference between Test 3 and PRISM observation. The test conditions are summarized in section 2.2.

The higher CAPE speeds up the development and increases the intensity of convection. A second reason is that Test 1 (K-F) generates 4–12% more precipitable water (PW, Figure 9a) than the control simulation (G-D) over the SC, which combines with the larger CAPE and triggers more convection. Third, to maintain moisture conservation in the model, larger amount of hydrometers aloft will require less amount of precipitation at the ground, and vice versa [*Hong and Lim*, 2006; *Hong et al*., 2009; *Cossu and Hocke*, 2014]. We compare the evolution of daily precipitation and hydrometer water path in JJA 2005 between Test 1 and the control simulation, defining the hydrometer water path as vertically integrated total condensates (in kg m^{-2}, including rain water, cloud water, snow, ice, and graupel). Test 1 (K-F) produces slightly greater precipitation than the control simulation (G-D) over the four subregions, with the largest differences (~15%) over SW (Figure 10a) and NR (Figure 10c). This is because there are fewer hydrometers (~30% less over SW, Figure 10b) aloft for Test 1 than for the control simulation.

We also compare the modeled atmospheric circulation between the control simulation and Test 1. The changes in circulation are partially attributable to the changes in latent heating associated with phase changes. Latent heating have been shown previously as an important energy source of large-scale atmospheric motions

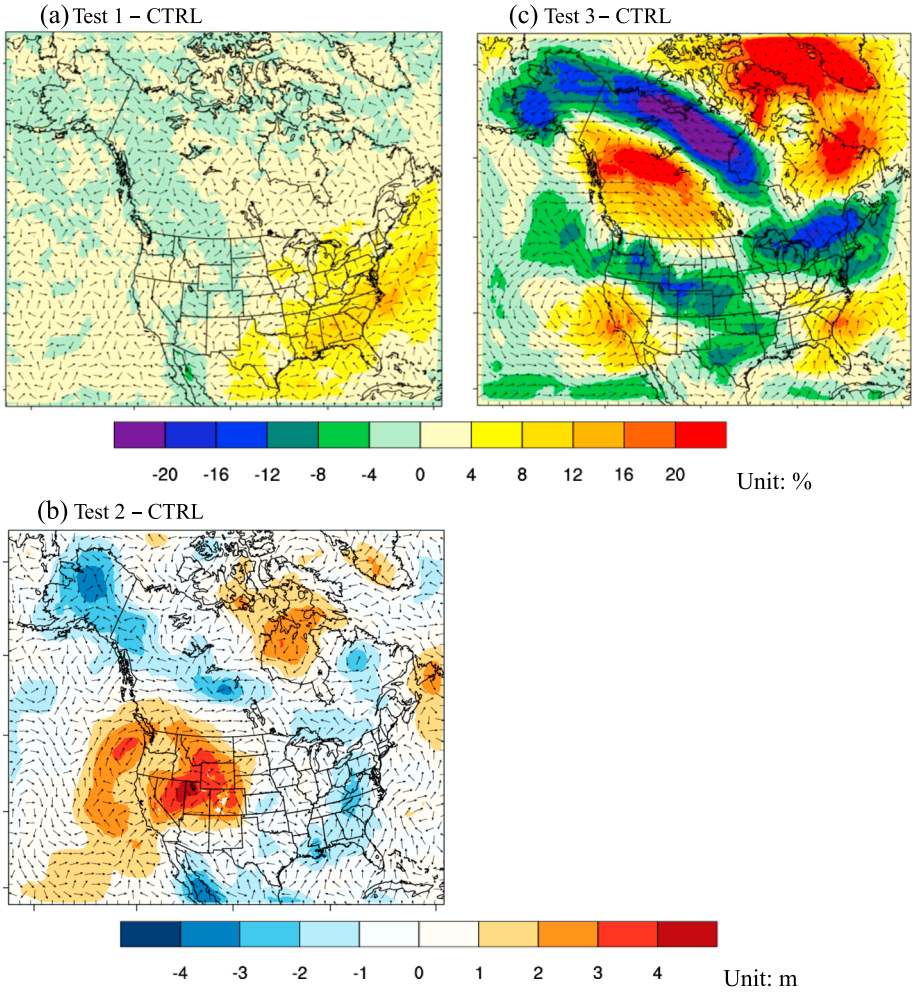

Figure 9. Precipitable water (shadings) and 700 hPa wind (vector) differences (a) between Test 1 and the control simulation in JJA 2005 and (c) between Test 3 and the control simulation in DJF 2004–2005. (b) Differences in 700 hPa geopotential height (shadings) and wind (vectors) between Test 2 and the control simulation in JJA. The range of wind speed difference is 0–1.5 m s^{-1} for Figure 9a and 0.5–2.0 m s^{-1} for Figures 9b and 9c.

[*Ling and Zhang*, 2013; *Sun et al.*, 2014]. However, we find no significant changes in low-level wind/geopotential height (Figure 9a) between the two simulations, because there is no significant change in vertical profile of latent heating.

The strong wet bias generated by Test 1 (K-F) could also be attributable to previously identified problems with the K-F scheme [*Tripathi and Dominguez*, 2013]. Although the cloud radius in K-F for WRF v3.3.1 is not constant, it is limited to the range 1–2 km, depending on the vertical velocity. This range might be too large for small-scale clouds. Because the mixing rate in the K-F scheme is inversely proportional to the subgrid scale updraft cloud radius, larger cloud radius leads to weaker mixing rate, which tends to strengthen the thermal contrast between the cloud and the environment, thus creating stronger convections [*Narita*, 2010; *Yang et al.*, 2012; *Tripathi and Dominguez*, 2013]. Additionally, as discussed in previous work [*Kain and Fritsch*, 1990, 1993; *Grell and Devenyi*, 2002; *Kain*, 2004; *Liang et al.*, 2004a], K-F and G-D schemes differ in their representations of the mechanism of precipitation production. The K-F scheme incorporates detailed cloud microphysics and entrainment/detrainment between the cloud and the environment. Its closure assumption is based on a CAPE removal process. In the K-F scheme, the updraft, downdraft, and environmental mass fluxes are rearranged in the mass column until 90% of the CAPE is removed. In contrast, the G-D scheme uses a simplified one-dimensional cloud model to calculate the cloud properties, in which the assumptions directly influence the vertical redistribution of heat/moisture or the surface precipitation rate. The G-D scheme employs

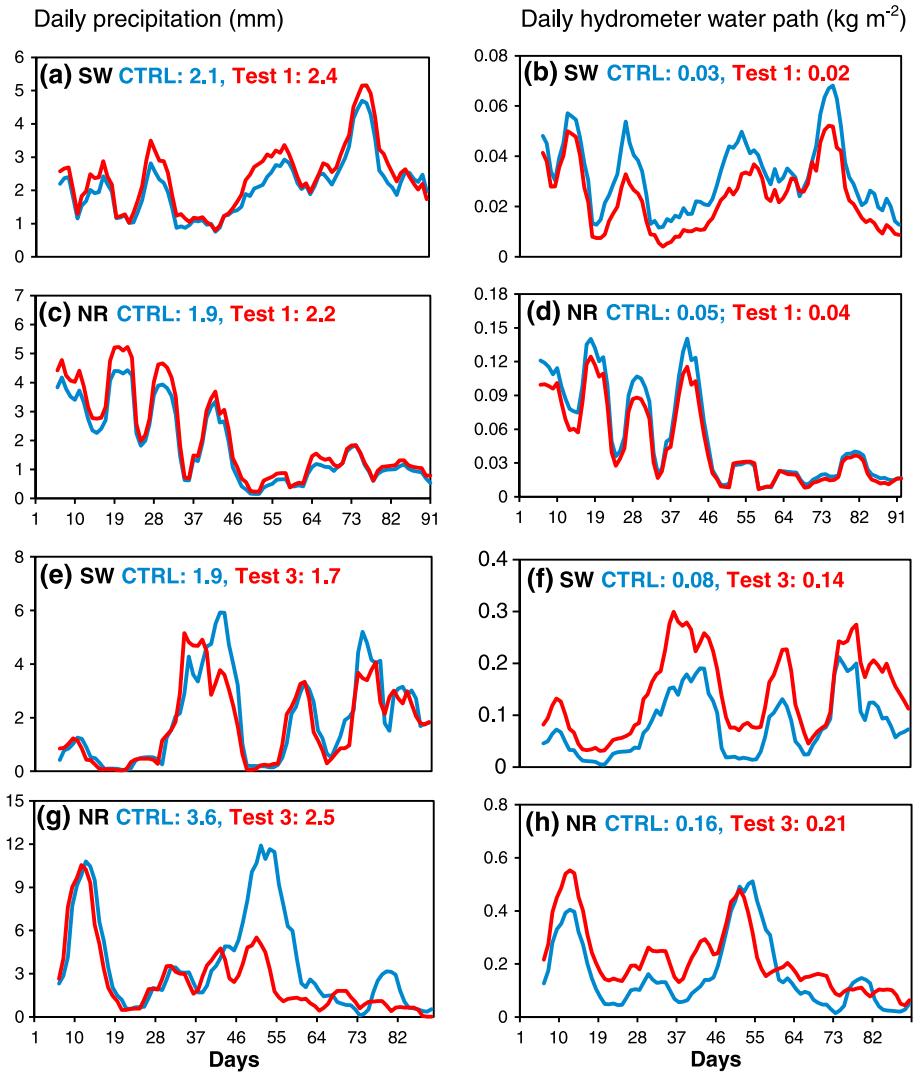

Figure 10. Times series of the surface precipitation (Figures 10a, 10c, 10e, and 10g) and hydrometer water path (Figures 10b, 10d, 10f, and 10h) over Southwest (SW) and North Rockies (NR) (a–d) for JJA 2005 resulting from the control simulation (blue) and Test 1 (red) and (e–h) for DJF 2004–2005 resulting from the control simulation (blue) and Test 3 (red). The numbers are seasonal means of precipitation and hydrometers over each subregion.

a variety of mass flux and closure assumptions, feedback assumptions, and trigger functions at each grid point. The closure assumptions include vertical advection of moisture, vertical velocity, and CAPE removal (similar to that in K-F). The rainfall is then parameterized by K-F with integrated water vapor and liquid flux above the lifting condensation level, which is parameterized by G-D with total condensate and cloud base mass flux in the updraft. Given the complexity of nonlinear interactions between various physics processes, it is extremely difficult to identify the individual differences likely to explain the WRF result differences for G-D versus K-F. The above analysis and discussion show how many small differences in two simulations can combine to produce relatively large overall impacts.

5.3. Nudging Strength

Figure 8c shows the rainfall differences between Test 2 with weaker nudging strength ($3 \times 10^{-5}\,s^{-1}$) and the PRISM observation in JJA 2005. Interestingly, the wet biases over western CONUS and GP are reduced by 1–2 mm d^{-1} in comparison with Figure 8a. Moreover, the dry biases around Kansas, Oklahoma, Missouri, and Arkansas are reduced by 1–3 mm d^{-1} versus Figure 8a. Both of these improvements make Test 2 closer than the control simulation to the observations. However, the precipitation over SC that is 1–2 mm d^{-1} heavier

than the observation needs to be investigated further. In addition, the warm bias over SW and SC in summer generated by weaker nudging is slightly stronger (~0.5°C, not shown) than the control simulation.

To investigate the impacts of nudging strength on atmospheric circulation, we compare the differences in wind and geopotential height at 850 and 700 hPa between Test 2 and the control simulation. As shown in Figure 9b, Test 2 (weaker nudging) generates higher geopotential heights and anticyclonic flow departures at 700 hPa (and 850 hPa) over northwestern CONUS. Meanwhile, the 500 hPa geopotential height over this area is 2.5 m higher than that of the control simulation (not shown). These circulation differences tend to reduce the precipitation over western CONUS. In contrast, Test 2 (weaker nudging) generates lower heights and southerly flow departures at 700 hPa (and 850 hPa) from the Gulf of Mexico to southeastern CONUS. The 500 hPa height over this area is also lower (0.5–1.5 m) than in the control simulation. These changes could increase the precipitation over SC.

5.4. Morrison Microphysics Scheme

Figures 8d compares PRISM precipitation values and WRF control simulation values in winter 2004–2005 (DJF). The precipitation bias is similar to the long-term winter means, with wet bias over northwestern CONUS and dry bias over the SC and most of California. Figure 8e shows the rainfall difference between Test 3 with the Morrison microphysics scheme and the PRISM values. The most significant improvement produced by Test 3 is reduction of the wet bias over western CONUS by up to 2–3 mm d^{-1}. This is important, because a large portion of annual precipitation and the extremes over the western United States occur mostly in the cold season [*Gutowski et al.*, 2008, 2010; *Dulière et al.*, 2011]. Understanding model bias over this region gives good insight into evaluating model performance for extreme events, especially given research findings that extreme precipitation is expected to increase with changing climate [*Kunkel et al.*, 1999; *Wehner*, 2005; *Tebaldi et al.*, 2006; *Kharin et al.*, 2007]. In addition, the wet biases in winter precipitation over GP and the Midwest are reduced by ~1 mm d^{-1}, making these results closer to the observations than the control simulation (Figures 3e, 3g, 5a, and 5c).

The reduced wet biases induced by Test 3 are mostly (>90%) contributed by the microphysics-resolved precipitation. Namely, Test 3 (Morrison) produces significantly less resolved precipitation and slightly more convective precipitation than the control simulation (WSM6). Note that >80% of winter precipitation is microphysics-resolved precipitation. One of the reasons for the reduced precipitation is that Test 3 (Morrison) produces less PW than the control simulation (WSM6) over western CONUS and southern GP, by 8–16% (0.5–1 mm) and 4–12% (0.5–1.5 mm), respectively. However, Test 3 (Morrison) generates greater PW than the control simulation over California, by 8–16% (1.5–2 mm). The lower PW over northwestern CONUS tends to generate less precipitation, while the higher PW over California and the southern Midwest could generate more precipitation. We also compare the evolution of daily precipitation and hydrometer water path in DJF 2004–2005 between Test 3 (Morrison) and the control simulation (WSM6). Test 3 (Morrison) produces significantly less precipitation than the control simulation (WSM6) over the four subregions, with the highest percentage difference (~30%) over NR (Figure 10g). This is because more hydrometers are aloft for Test 3 than for the control simulation (Figure 10h).

The atmospheric circulations modeled by Test 3 (Morrison) and the control simulation are different, partly because of the changes in latent heating. In comparison with the control simulation, Test 3 (Morrison) reduces the latent heating rate by up to 3 K d^{-1} over SW and NR and by 1–2 K d^{-1} over GP. Meanwhile, Test 3 (Morrison) increases the latent heating rate by 2–3 K d^{-1} over California. These differences mainly occur below 500 hPa. Accordingly, as shown in Figure 9c, anticyclonic flow departures generated by Test 3 over northwestern CONUS brings the dry/cold air from the north and prevent the development of precipitation over the Northwest and the northern GP. In contrast, cyclonic flow departures over the Pacific Ocean and California bring moisture from the Pacific to California, tending to produce more precipitation than in the control simulation.

The most important difference between the WSM6 and Morrison scheme is that the WSM6 assumes a constant size distribution intercept parameter for each precipitation species except snow. In the Morrison scheme, the size distribution intercept parameter evolves freely from the predicted number concentration and mixing ratio, allowing for more flexible treatment of the particle size distributions [*Morrison et al.*, 2009]. However, other differences also exist between the two schemes. For example, as stated by *Liu et al.* [2011], in comparison with Morrison scheme, the WSM6 initiates a faster conversion from vapor to ice crystals when ice supersaturation is achieved. The ice crystal is much more prone to fall as precipitation

Table 1. Control- and Test 4-Simulated Regional-Average Bias for Temperature in 2005[a]

	Temperature Bias (°C)							
	GP		SW		SC		NR	
	CTRL	Test 4	CTRL	Test 4	CTRL	Test 4	CTRL	Test 4
Spring	1.98	0.45	0.05	**0.08**	2.16	1.77	0.34	0.31
Summer	1.38	**1.61**	0.71	**1.96**	0.61	0.37	−0.83	0.10
Fall	2.04	**2.94**	0.37	**1.77**	2.35	**2.47**	0.93	**1.66**
Winter	1.54	**2.78**	−0.59	**1.11**	3.92	3.82	2.16	**4.23**

[a]Bold font indicates a larger bias for Test 4 than for the control simulation.

than are cloud droplets, possibly leading to a relatively high precipitation efficiency (defined as the ratio of the surface precipitation to the hydrometer water path). On the other hand, the Morrison scheme is more detailed and critical in representing each graupel production pathway than is WSM6. The Morrison scheme converts only a partial amount of combined snow and cloud water into graupel, while 5–75% of the mass remains in the original category. In contrast, in WSM6, the efficiency of snow-collecting cloud water may be nearly 1, which is not always the case in observations.

5.5. Longer Spin-Up Time

Tables 1 and 2 compare subregional-average biases versus the PRISM observations for temperature and precipitation in all four seasons from the WRF control simulation with 1 day spin-up time and Test 4 with 1 year spin-up time. As Table 1 shows, the significant warm bias over SC (Figure 3c) simulated by the control is reduced by Test 4 from 2.16, 0.61, and 3.92°C to 1.77, 0.37, and 3.82°C in spring, summer, and winter, respectively. The warm biases over GP and NR in spring are also reduced by Test 4 from 1.98 and 0.34°C to 0.45 and 0.31°C, respectively. The cold bias (−0.83°C) over NR in summer is largely eliminated by Test 4 (0.1°C). However, Test 4 generates a stronger warm bias than the control simulation over all subregions in fall and winter, except over SC in winter. For example, the warm biases in fall are increased by Test 4 from 2.04, 0.37, 2.35, and 0.93°C to 2.94, 1.77, 2.47, and 1.66°C over GP, SW, SC, and NR, respectively. Moreover, the warm bias over GP and SW in summer is increased by Test 4 from 1.38 and 0.71°C to 1.61 and 1.96°C, respectively. Figure S2 presents the spatial patterns of temperature for the PRISM observations, the control simulation, and Test 4 in four seasons.

As Table 2 shows, the significant wet bias over GP (Figure 3e) in the control simulation is reduced by Test 4 from 1.44, 1.14, and 0.57 mm d^{-1} to 1.02, 0.07, and 0.18 mm d^{-1} in spring, summer, and fall, respectively. The wet bias over SW is also reduced from 3.65, 4.2, 1.23, and 1.5 mm d^{-1} to 0.7, 1.58, 0.27, and 0.28 mm d^{-1} in spring, summer, fall and winter, respectively. However, Test 4 shows stronger dry bias over eastern CONUS in fall and winter than the control simulation. The absolute bias over SC is increased by Test 4 from 0.12 and 0.11 mm d^{-1} to 0.58 and 0.43 mm d^{-1} in fall and winter, respectively. In addition, the precipitation over NR is overestimated more by Test 4 than by the control simulation, and the wet bias is increased from 1.55, 0.89, and 1.45 mm d^{-1} to 1.7, 1.28, and 2.61 mm d^{-1} in spring, fall, and winter, respectively. Again, these biases affect the model performance in extreme precipitation over this region, which occurs mostly in the cold season. Figure S3 presents the spatial patterns of precipitation from the PRISM values, the control simulation, and Test 4 in four seasons.

Table 2. Control- and Test 4-Simulated Regional-Average Bias for Precipitation in 2005[a]

	Precipitation Bias (mm d^{-1})							
	GP		SW		SC		NR	
	CTRL	Test 4	CTRL	Test 4	CTRL	Test 4	CTRL	Test 4
Spring	1.44	1.02	3.65	0.70	1.61	1.08	1.55	**1.70**
Summer	1.14	0.07	4.20	1.58	1.70	**3.13**	0.51	0.35
Fall	0.57	0.18	1.23	0.27	−0.12	**−0.58**	0.89	**1.28**
Winter	0.62	**0.77**	1.50	0.28	−0.11	**−0.43**	1.45	**2.61**

[a]Bold font indicates a larger bias for Test 4 than for the control simulation.

6. Summary and Discussion

6.1. Summary

The performance of a 12 km WRF downscaling in near-surface air temperature and precipitation driven by NCEP-R2 reanalysis data over CONUS in 1980–2004 is assessed on various temporal and spatial scales by using high-resolution data sets. The WRF shows reasonable downscaling skills for air temperature and precipitation, producing realistic geographic distributions and seasonal variations. The WRF performs better than the 50 km NARCCAP-WRFG over different subregions in different seasons, and it also adds value above the NCEP-R2, especially over complex terrain (e.g., mountains and coastal zones).

However, the WRF-simulated temperature over SC and precipitation over GP and SW remain problematic. Some of the model biases lead to poorer performance than the coarse-resolution simulation (NARCCAP-WRFG) and the driving data (NCEP-R2). Spectral nudging strength, spin-up time, integration method (reinitialization versus continuous integration), and microphysics scheme have important effects on the calculated precipitation and temperature, but the cumulus parameterizations tested do not show significant effect. Reducing the spectral nudging strength and/or allowing longer spin-up time can partly address the wet bias over GP and SW and the warm bias over SC, but these changes generate larger bias for precipitation over SC and NR and for temperature over GP and SW. Replacing the WSM6 microphysics scheme with the Morrison scheme reduces the wet (dry) bias over the western mountain ranges and GP (over California) in winter.

6.2. Discussion

In analyzing the precipitation bias in the WRF model, *Wang and Kotamarthi* [2013] found that turning off the spectral nudging substantially reduced the summer rainfall over GP. However, it generated $>$100% dry bias over Kansas and Oklahoma. On the other hand, the wet bias over GP and northwestern CONUS is likely due to the cumulus parameterization, since we find that both the G-D and K-F cumulus parameterizations generate significant wet bias over those regions (Figures 8a and 8b). A simulation at 4 km grid resolution for the summer of 2005, with the cumulus parameterization turned off, is shown to reduce the wet bias over northern GP by ~30% in comparison with the control simulation using G-D. The biases in the WRF simulation could also result from uncertainties in the lateral boundary condition (LBCs). For example, *Liang et al*. [2004a] found that using the European Center for Medium-Range Weather Forecast reanalysis can address the winter dry biases over SC, which likely result from LBC errors for NCEP-R2 data.

The differences in the physics parameterizations and model configurations between the WRF and the NARCCAP-WRFG make a direct comparison difficult. The improvements of the WRF over the NARCCAP-WRFG shown in sections 3 and 4 result from the combined effects of these differences. To make the comparisons structured and illustrative, we evaluate the effect of each of these differences between the models and identify the likely combination of changes that leads to the net calculated difference between the calculations. *Miguez-Macho et al*. [2004] showed that spectral nudging successfully eliminated the differences in climate variables even when the model domain was shifted by 7–17°. *Liu et al*. [2011] found that the WRF model performance was not significantly different when NCAR Community Atmosphere Model radiation and the RRTM scheme were used. Therefore, the differences in model domain sizes and radiation schemes are probably less important to the differences between WRF and NARCCAP-WRFG than other effects. We investigate the effect of model resolution on model performance by running a 1 year, 50 km simulation over the same domain and under the same conditions as the 12 km simulation. The result shows that the fine-resolution simulation indeed adds detailed value/processes to the coarse-resolution simulation, especially over complex topographies (e.g., mountain ranges and coastal zones). For example, the fine-resolution simulation reduces the absolute (cold) bias over SW and NR from 1.6–3.6°C to 0.7–1.2°C and the wet bias over SW and NR from 0.5–1.6 mm to 0.4–1.3 mm (annual average). The reinitializing integration method, which increases the observation information and reduces possible drifts caused by accumulated model errors, is the other reason for the improvements. As illustrated by Test 4 (section 5.5 and Table 2), the drier bias over SC (Figure 3g) and the wetter bias over NR (Figure 3h) in the NARCCAP-WRFG than in the WRF are most likely induced by the continuous integration of the model. Spectral nudging does increase the skill of a regional climate model [e.g., *von Storch et al*., 2000; *Laprise et al*., 2003; *Miguez-Macho et al*., 2005; *Castro et al*., 2005; *Laprise et al*., 2008; *Bowden et al*., 2012; *Otte et al*., 2012] in capturing the climatological features of atmospheric variables. In this study, a 1 year simulation at 50 km without spectral nudging misses the rainfall band over the GP in summer and fall while showing a wetter bias in winter over SW and NR than the 50 km

simulation with spectral nudging. The simulation without spectral nudging produces much weaker low-level jet from the Gulf of Mexico to the GP (Figure S4), reducing moisture transport from the Gulf of Mexico to the GP and tending to reduce precipitation in summer [*Liang et al*., 2004a; *Martynov et al*., 2013]. Moreover, the simulation without spectral nudging shows higher 500 mb height, by up to 12.5 m, over GP and eastern CONUS than the simulation with spectral nudging. Meanwhile, it shows lower 500 mb height, by 5–6 m, over western CONUS (Figure S5). These differences explain the dry bias over GP and eastern CONUS and wet bias over western CONUS generated by the simulation without spectral nudging.

Acknowledgments

We thank all anonymous reviewers and Dr. Virendra P. Ghate at Argonne for their constructive comments and insights. This work is supported under a military interdepartmental purchase request from the Strategic Environmental Research and Development Program, RC-2242, through U.S. Department of Energy (DOE) contract DE-AC02-06CH11357. We acknowledge the PRISM group (http://www.prism.oregonstate.edu/), the NARCCAP group (http://www.narccap.ucar.edu/about/index.html), and the NOAA/OAR/ESRL PSD (ftp.cdc.noaa.gov/Datasets/NARR/monolevel/) for providing the observed monthly data sets, the RCM outputs, and the NARR 3 h precipitation data, respectively. The computational resources are provided by the DOE-supported Argonne Leadership Computing Facility and the National Energy Research Scientific Computing Center.

References

Bowden, J. H., T. L. Otte, C. G. Nolte, and M. J. Otte (2012), Examining interior grid nudging techniques using two-way nesting in the WRF model for regional climate modeling, *J. Clim.*, *25*, 2805–2823.

Bukovsky, M. S., and D. J. Karoly (2007), A brief evaluation of precipitation from the North American regional reanalysis, *J. Hydrometeorol.*, *8*, 837–846.

Castro, C. L., R. A. Pielke Sr., and G. Leoncini (2005), Dynamical downscaling: Assessment of value retained and added using the Regional Atmospheric Modeling System (RAMS), *J. Geophys. Res.*, *110*, D05108, doi:10.1029/2004JD004721.

Castro, C. L., R. A. Pielke Sr., and J. O. Adegoke (2007), Investigation of the summer climate of the contiguous U.S. and Mexico using the regional atmospheric modeling system (RAMS). Part I: Model climatology (1950–2002), *J. Clim.*, *20*, 3866–3887.

Chen, F., and J. Dudhia (2001), Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity, *Mon. Weather Rev.*, *129*, 569–585.

Conil, S., and A. Hall (2006), Local regimes of atmospheric variability: A case study of southern California, *J. Clim.*, *19*, 4308–4325.

Cossu, F., and K. Hocke (2014), Influence of microphysical schemes on atmospheric water in the Weather Research and Forecasting model, *Geosci. Model Dev.*, *7*, 147–160, doi:10.5194/gmd-7-147-2014.

Daly, C., R. P. Neilson, and D. L. Phillips (1994), A statistical-topographic model for mapping climatological precipitation over mountainous terrain, *J. Appl. Meteorol.*, *33*, 140–158.

Daly, C., G. H. Taylor, and W. P. Gibson (1997), The PRISM approach to mapping precipitation and temperature, Preprints, paper presented at 10th Conference on Applied Climatology, Am. Meteorol. Soc., Reno, Nev.

Daly, C., M. Halbleib, J. I. Smith, W. P. Gibson, M. K. Doggett, G. H. Taylor, J. Curtis, and P. P. Pasteris (2008), Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, *Int. J. Climatol.*, *28*, 2031–2064.

de Ela, R., R. Laprise, and B. Denis (2002), Forecasting skill limits of nested, limited-area models: A perfect model approach, *Mon. Weather Rev.*, *130*, 2006–2023.

Denis, B., R. Laprise, D. Caya, and J. Côté (2002), Downscaling ability of one-way nested regional climate models: The Big-Brother Experiment, *Clim. Dyn.*, *18*, 627–646, doi:10.1007/s00382-001-0201-0.

Dulière, V., Y. Zhang, and E. P. Salathé (2011), Extreme precipitation and temperature over the U.S. Pacific Northwest: A comparison between observations, reanalysis data, and regional models, *J. Clim.*, *24*, 1950–1964.

Feser, F. (2006), Enhanced detectability of added value in limited-area model results separated into different spatial scales, *Mon. Weather Rev.*, *134*, 2180–2190, doi:10.1175/MWR3183.1.

Giorgi, F., and L. O. Mearns (1999), Introduction to special section: Regional climate modeling revisited, *J. Geophys. Res.*, *104*, 6335–6352, doi:10.1029/98JD02072.

Giorgi, F., C. Jones, and G. R. Asrar (2009), Addressing climate information needs at the regional level: The CORDEX framework, *WMO Bull.*, *58*, 175–183.

Grell, G. A., and D. Devenyi (2002), A generalized approach to parameterizing convection combining ensemble and data assimilation techniques, *Geophys. Res. Lett.*, *29*(14), 1693, doi:10.1029/2002GL015311.

Gutowski, W. J., S. S. Willis, J. C. Patton, B. R. J. Schwedler, R. W. Arritt, and E. S. Takle (2008), Changes in extreme, cold-season synoptic precipitation events under global warming, *Geophys. Res. Lett.*, *35*, L20710, doi:10.1029/2008GL035516.

Gutowski, W. J., et al. (2010), Regional extreme monthly precipitation simulated by NARCCAP RCMs, *J. Hydrometeorol.*, *11*, 1373–1379.

Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister (2013), Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 Dataset, *Int. J. Climatol.*, doi:10.1002/joc.3711.

Hong, S. Y., and J. J. Lim (2006), The WRF single-moment 6-class microphysics scheme (WSM6), *J. Korean Meteor. Soc.*, *42*, 129–151.

Hong, S. Y., K.-S. S. Lim, J.-H. Kim, J.-O. J. Lim, and J. Dudhia (2009), Sensitivity study of cloud-resolving convective simulations with WRF using two bulk microphysical parameterizations: Ice-phase microphysics versus sedimentation effects, *J. Appl. Meteorol. Climatol.*, *48*, 61–76, doi:10.1175/2008JAMC1960.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins (2008), Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models, *J. Geophys. Res.*, *113*, D13103, doi:10.1029/2008JD009944.

Jiménez, P. A., J. F. González-Rouco, E. García-Bustamante, J. Navarro, J. P. Montávez, J. V. de Arellano, J. Dudhia, and A. Muñoz-Roldan (2010), Surface wind regionalization over complex terrain: Evaluation and analysis of a high-resolution WRF simulation, *J. Appl. Meteorol. Climatol.*, *49*, 268–287, doi:10.1175/2009JAMC2175.1.

Kain, J. S. (2004), The Kain–Fritsch convective parameterization: An update, *J. Appl. Meteorol.*, *43*, 170–181.

Kain, J. S., and J. M. Fritsch (1990), A one-dimensional entraining/detraining plume model and its application in convective parameterization, *J. Atmos. Sci.*, *47*, 2784–2802.

Kain, J. S., and J. M. Fritsch (1993), Convective parameterization for mesoscale models: The Kain-Fritsch scheme. The representation of cumulus convection in numerical models, *Meteorol. Monogr.*, *24*, 165–170.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter (2002), NCEP–DOE AMIP-II Reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, *83*, 1631–1643.

Kharin, V., F. W. Zwiers, X. Zhang, and G. C. Hegerl (2007), Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, *J. Clim.*, *20*, 1419–1444.

Kunkel, K. E., K. Andsager, and D. R. Easterling (1999), Long-term trends in extreme precipitation events over the conterminous United States and Canada, *J. Clim.*, *12*, 2515–2527.

Laprise, R., D. Caya, A. Frigon, and D. Paquin (2003), Current and perturbed climate as simulated by the second-generation Canadian Regional Climate Model (CRCM-II) over northwestern North America, *Clim. Dyn.*, *21*, 405–21.

Laprise, R., et al. (2008), Challenging some tenets of regional climate modelling, *Meteorol. Atmos. Phys.*, *100*, 3–22, doi:10.1007/s00703-008-0292-9.
Lebassi-Habtezion, B., and N. S. Diffenbaugh (2013), Nonhydrostatic nested climate modeling: A case study of the 2010 summer season over the western United States, *J. Geophys. Res. Atmos.*, *118*, 10,944–10,962, doi:10.1002/jgrd.50773.
Liang, X. Z., L. Li, K. E. Kunkel, M. Ting, and J. Wang (2004a), Regional climate model simulation of U.S. precipitation during 1982–2002. Part I: Annual cycle, *J. Clim.*, *17*, 3510–3529.
Liang, X. Z., L. Li, A. Dai, and K. E. Kunkel (2004b), Regional climate model simulation of summer precipitation diurnal cycle over the United States, *Geophys. Res. Lett.*, *31*, L24208, doi:10.1029/2004GL021054.
Ling, J., and C. Zhang (2013), Diabatic heating profiles in recent global reanalyses, *J. Clim.*, *26*, 3307–3325, doi:10.1175/JCLI-D-12-00384.1.
Liu, C., K. Ikeda, G. Thompson, R. Rasmussen, and J. Dudhia (2011), High-resolution simulations of wintertime precipitation in the Colorado headwaters region: Sensitivity to physics parameterizations, *Mon. Weather Rev.*, *139*, 3533–3553, doi:10.1175/MWR-D-11-00009.1.
Lo, J. C., Z. L. Yang, and R. A. Pielke Sr. (2008), Assessment of three dynamical climate downscaling methods using the Weather Research and Forecasting (WRF) model, *J. Geophys. Res.*, *113*, D09112, doi:10.1029/2007JD009216.
Lucas-Picher, P., F. Boberg, J. H. Christensen, and P. Berg (2013), Dynamical downscaling with reinitializations: A method to generate finescale climate datasets suitable for impact studies, *J. Hydrometeorol.*, *14*, 1159–1174, doi:10.1175/JHM-D-12-063.1.
Martynov, A., R. Laprise, L. Sushama, K. Winger, L. Šeparović, and B. Dugas (2013), Reanalysis-driven climate simulation over CORDEX North America domain using the Canadian Regional Climate Model, version 5: Model performance evaluation, *Clim. Dyn.*, *41*, 2973–3005, doi:10.1007/s00382-013-1778-9.
Matsuura, K., and C. Willmott (2010), Terrestrial air temperature and precipitation: 900-2008 gridded monthly time series (V 2.01), University of Delaware Department of Geography Center for Climatic Research. [Available online at http://climate.geog.udel.edu/~climate/html_pages/archive.html.]
Mearns, L. O., et al. (2007), Updated 2012, The North American Regional Climate Change Assessment Program dataset, National Center for Atmospheric Research Earth System Grid data portal, Boulder, Colo. Data downloaded 2013-03-11, doi:10.5065/D6RN35ST.
Mearns, L. O., W. Gutowski, R. Jones, R. Leung, S. McGinnis, A. Nunes, and Y. Qian (2009), A regional climate change assessment program for North America, *EOS Trans.*, *90*, 311–311, doi:10.1029/2009EO360002.
Mearns, L. O., et al. (2012), The North American Regional Climate Change Assessment program: Overview of phase I results, *Bull. Am. Meteorol. Soc.*, *93*, 1337–1362.
Mesinger, F., et al. (2006), North American regional reanalysis, *Bull. Am. Meteorol. Soc.*, *87*, 343–360.
Miguez-Macho, G., G. L. Stenchikov, and A. Robock (2004), Spectral nudging to eliminate the effects of domain position and geometry in regional climate model simulations, *J. Geophys. Res.*, *109*, D13104, doi:10.1029/2003JD004495.
Miguez-Macho, G., G. L. Stenchikov, and A. Robock (2005), Regional climate simulations over North America: Interaction of local processes with improved large-scale flow, *J. Clim.*, *18*, 1227–1246.
Mitchell, T. D., and P. D. Jones (2005), An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *Int. J. Climatol.*, *25*, 693–712.
Morrison, H., G. Thompson, and V. Tatarskii (2009), Impact of cloud microphysics on the development of trailing stratiform precipitation in a simulated squall line: Comparison of one- and two-moment schemes, *Mon. Weather Rev.*, *137*, 991–1007, doi:10.1175/2008MWR2556.1.
Narita, M. (2010), Modification to the mixing rate of convective cloud in the Kain-Fritsch scheme, CAS/JSC WGNE Res. Activ. Atmos. Oceanic Modell., 40, Section 04, 11–12.
Noh, Y., W. G. Cheon, S. Y. Hong, and S. Raasch (2003), Improvement of the K-profile model for the planetary boundary layer based on large eddy simulation data, *Bound. Layer Meteor.*, *107*, 401–427.
Omrani, H., P. Drobinski, and T. Dubos (2012), Spectral nudging in regional climate modelling: How strongly should we nudge?, *Q. J. R. Meteorol. Soc.*, *138*, 1808–1813, doi:10.1002/qj.1894.
Omrani, H., P. Drobinski, and T. Dubos (2013), Optimal nudging strategies in regional climate modelling: Investigation in a Big-Brother experiment over the European and Mediterranean regions, *Clim. Dyn.*, *41*, 2451–2470.
Otte, T. L., C. G. Nolte, M. J. Otte, and J. H. Bowden (2012), Does nudging squelch the extremes in regional climate modeling?, *J. Clim.*, *25*, 7046–7066.
Pan, Z., E. Takle, W. Gutowski, and R. Turner (1999), Long simulation of regional climate as a sequence of short segments, *Mon. Weather Rev.*, *127*, 308–321.
Pielke, R. A., Sr. (2002), *Mesoscale Meteorological Modeling*, *Int'l. Geophys. Ser.*, vol. 78, 2nd ed., 676 pp., Academic Press, San Diego, Calif.
Qian, J.-H., A. Seth, and S. Zebiak (2003), Reinitialized versus continuous simulations for regional climate downscaling, *Mon. Weather Rev.*, *131*, 2857–2874.
Rockel, B., C. L. Castro, R. A. Pielke Sr., H. von Storch, and G. Leoncini (2008), Dynamical downscaling: Assessment of model system dependent retained and added variability for two different regional climate models, *J. Geophys. Res.*, *113*, D21107, doi:10.1029/2007JD009461.
Rummukainen, M. (2010), State-of-the-art with regional climate model, *Wiley Interdiscip. Rev. Clim. Change*, *1*, 82–96, doi:10.1002/wcc.8.
Salzmann, N., and L. O. Mearns (2012), Assessing the performance of multiple regional climate model simulations for seasonal mountain snow in the upper Colorado River basin, *J. Hydrometeorol.*, *13*, 539–556.
Sun, Y., Z. Zhong, W. Lu, and Y. Hu (2014), Why are tropical cyclone tracks over the Western North Pacific sensitive to the cumulus parameterization scheme in regional climate modeling? A case study for Megi (2010), *Mon. Weather Rev.*, *142*, 1240–1249.
Tebaldi, C., K. Hayhoe, J. M. Arblaster, and G. A. Meehl (2006), Going to the extremes: An intercomparison of model-simulated historical and future changes in extreme events, *Clim. Change*, *79*, 185–211.
Tian, B., I. M. Held, N.-C. Lau, and B. J. Soden (2005), Diurnal cycle of summertime deep convection over North America: A satellite perspective, *J. Geophys. Res.*, *110*, D08108, doi:10.1029/2004JD005275.
Tripathi, O. P., and F. Dominguez (2013), Effects of spatial resolution in the simulation of daily and subdaily precipitation in the southwestern US, *J. Geophys. Res. Atmos.*, *118*, 7591–7605, doi:10.1002/jgrd.50590.
von Storch, H., H. Langenberg, and F. Feser (2000), A spectral nudging technique for dynamical downscaling purposes, *Mon. Weather Rev.*, *128*, 3664–73.
Wang, J., and V. R. Kotamarthi (2013), Assessment of dynamical downscaling in near-surface fields with different spectral nudging approaches using the Nested Regional Climate Model (NRCM), *J. Appl. Meteor. Climatol.*, *52*, 1576–1591.
Wehner, M. F. (2005), Changes in daily precipitation and surface air temperature extremes in the IPCC AR4 models, *US CLIVAR Variations*, *3*, 5–9.

Wi, S., F. Dominguez, M. Durcik, J. B. Valdes, H. F. Diaz, and C. L. Castro (2012), Climate change projection of snowfall in the Colorado River Basin using dynamical downscaling, *Water Resour. Res.*, *48*, W05504, doi:10.1029/2011WR010674.

Willmott, C. J., and K. Matsuura (1995), Smart interpolation of annually averaged air temperature in the United States, *J. Appl. Meteorol.*, *34*, 2577–2586.

Yang, B., Y. Qian, G. Lin, R. Leung, and Y. Zhang (2012), Some issues in uncertainty quantification and parameter tuning: A case study of convective parameterization scheme in the WRF regional climate model, *Atmos. Chem. Phys.*, *12*, 2409–2427, doi:10.5194/acp-12-2409-2012.

Zagar, N., M. Zagar, J. Cedilnik, G. Gregoric, and J. Rakovec (2006), Validation of mesoscale low-level winds obtained by dynamical downscaling of ERA40 over complex terrain, *Tellus*, *58A*, 445–455.